

Appel à communications
Colloque interdisciplinaire – UPJV (Amiens)
10-11 mai 2012

« Complémentarité des approches qualitatives et quantitatives dans l'analyse des discours ? »

Organisé par le CURAPP (Centre universitaire de recherches sur l'action publique et le politique – Epistémologie et sciences sociales, UMR 6054)
Avec le soutien de l'Université de Picardie Jules Verne (UPJV)

Comité scientifique :

Marie-Anne Paveau (Université Paris 13) – Damon Mayaffre (Université de Nice) – Johannes Angermüller (Université de Mainz) – Jacques Guilhaumou (Université de Provence), Corinne Gobin (Université libre de Bruxelles) – Frédéric Lebaron (Université de Picardie)

Responsable du projet : Thierry Guilbert, maître de conférence en sciences du langage (analyse du discours) à l'UPJV, responsable du sous-axe : « Discours, identités, représentations partagées » (Curapp)
Co-organisatrice : Pascaline Lefort, docteure en sciences du langage (analyse du discours) (Curapp)

Contact : thierry.guilbert@u-picardie.fr

Ce colloque a pour objectif de réfléchir de façon collective et interdisciplinaire à l'articulation des approches quantitatives et qualitatives dans l'étude des discours en partant de ce simple constat : si l'étude des discours n'a jamais été l'apanage de l'analyse du discours (AD), aujourd'hui les études quantitatives sur le discours, hors de l'AD, se multiplient, notamment grâce à la diffusion, et à la qualité, des outils informatiques. Si bien que l'on pourrait, sans provocation gratuite, se poser la question suivante : que reste-t-il de la spécificité de l'AD aujourd'hui ?

La coexistence des deux approches qualitatives et quantitatives n'est pas nouvelle en AD. Depuis les années 60, celles-ci se sont développées conjointement, l'une semblant être le prolongement « naturel » de l'autre. Ainsi au moment où se constituaient les « linguistiques de la parole » comme a pu les appeler un moment Pêcheux, c'est-à-dire les approches pragmatico-énonciatives sur le langage et son utilisation (à partir des travaux de Benveniste et de Culioli notamment), celui-ci développait un dispositif critique d'analyse automatique du discours (AAD) à partir de la psychologie sociale¹ (Léon 2010). A la même époque, l'équipe de l'ENS de Saint-Cloud développait, autour de Tournier et Dubois, une analyse quantitative du vocabulaire politique en créant ses propres outils lexicométriques. Ensuite les outils se sont diversifiés en de nombreux travaux et logiciels quantitatifs (Muller, Tournier, Lafon, Lebart et Salem, Brunet) : lemmatiseurs, analyseurs syntaxiques, calculs statistiques, analyses des contextes et des spécificités, etc. (*Cordial*, *Winbrill*, *Alceste*, *Hyperbase*, *Tropes*, *Lexico*...)

¹ Critiques du modèle documentaire existant qu'un psychologue social comme R. Ghiglione a largement poursuivi par une méthode d'analyse propositionnelle.

Depuis quelques années, les logiciels qui permettent au chercheur d'être assisté par un ordinateur, que ce soit pour les décomptes d'occurrences ou pour les calculs statistiques, ont gagné en notoriété et se sont répandus hors du cadre des sciences du langage. Il est vrai que l'« analyse de textes assistée par ordinateur » présente un double intérêt, l'ordinateur est à la fois un outil probatoire et un outil heuristique, c'est-à-dire qu'il permet à la fois de fournir des preuves objectives et quantifiées et de proposer des pistes de réflexion au chercheur en lui faisant découvrir ce qu'il ne cherchait pas forcément (Mayaffre 2002). Cependant en débordant très largement du cadre de l'AD et du Traitement Automatique des Langues (TAL), l'utilisation de ces logiciels s'éloigne quelquefois également des présupposés épistémologiques de l'AD.

Ainsi, l'approche méthodologique consistant à compléter l'analyse quantitative par une analyse qualitative – voire à faire « fi du quantitatif » (Mazière *et al.* 2010) –, approche très répandue en analyse du discours, ne va pas forcément de soi pour les spécialistes en psychologie sociale, en sociologie, en sciences politiques ou encore en anthropologie qui analysent le langage dans le but d'y lire les pratiques des acteurs et les représentations des sujets. Parfois utilisé sans précautions méthodologiques – sans se soucier des présupposés épistémologiques des sciences du langage et de l'AD dans lequel il a pourtant été conçu –, le logiciel semble acquérir une sorte de vertu magique : il dirait de lui-même et par lui-même de quoi est fait le discours. On retrouve alors « l'éternel malentendu » entre AD et sciences sociales – « une conception transparente de la langue et du discours encore si prévalente dans les sciences humaines aujourd'hui » (Robin 1986, 128) – et, par certains aspects, les motifs de la critique des « analyses de contenu » de Pêcheux (1969), mais à nouveaux frais.

Ce qui est en jeu – il ne s'agit ni de généraliser abusivement un constat parcellaire, ni de stigmatiser une discipline particulière –, c'est le risque d'assister peu à peu à un *recul épistémologique* : des avancées depuis longtemps actées comme, par exemple, les réflexions sur la constitution des corpus (voir Mazière 2005), la remise en question de la bi-univocité du langage, la signification en contexte, la distinction langue/discours, voire la notion saussurienne de valeur du signe, semblent quelquefois proprement ignorées par certains travaux qui étudient le discours. Ainsi assiste-t-on parfois à une méconnaissance des « principes pour l'arbitrage des cas douteux »² (Muller 1973, 9), voire à l'*ignorance* de l'existence de ces cas douteux.

On pourrait objecter qu'à l'inverse, les chercheurs en sciences sociales (sociologues, politistes et psychologues notamment) sont souvent très bien formés à la statistique et maîtrisent les précautions qui président aux calculs statistiques (hypothèse nulle, loi normale, etc.), formation que n'ont pas forcément tous les analystes et les linguistes. Cela semble cependant moins dramatique d'un point de vue épistémologique car les logiciels prennent en compte dans leur programmation et dans leurs processus d'utilisation ces présupposés méthodologiques (voir *Hyperbase*, par exemple). Si on peut se féliciter de cet intérêt grandissant pour les discours et de l'élargissement du cercle des utilisateurs des logiciels, on ne peut que s'inquiéter de certaines de ces utilisations qui peuvent apparaître comme des dérives, une réflexion interdisciplinaire sur l'articulation entre les approches quantitatives et qualitatives semble alors nécessaire.

² Par exemple : « grand » classé uniquement comme un adjectif évaluatif alors qu'il peut être axiologique. Quelle validité accorder à des études qui se basent sur ce genre de « faits linguistiques » ?

Les contributions pourront ainsi traiter des questions suivantes (sans rechercher l'exhaustivité) :

1. Controverses et recherche de solutions :
 - a. Peut-on vraiment parler d'« utilisation sauvage » de ces logiciels ?
 - b. L'AD a-t-elle à prendre parti vis-à-vis de ces formes d'utilisation ? Si l'analyse du discours ne peut bien sûr s'ériger comme une norme d'utilisation, ou pire comme le garant de la « bonne » utilisation des logiciels, quelle position doit-elle prendre ?
 - c. L'analyse du discours est-elle cumulative ? Si oui, que faire de ce recul épistémologique ? Comment (re)mettre en lumière les réflexions épistémologiques qui président à l'utilisation de ces outils et à la fabrication des corpus ?
 - d. L'AD a-t-elle les moyens (et le droit) de revendiquer une interdisciplinarité des travaux ?
2. Questions méthodologiques :
 - a. Quid du corpus ? Et du rôle du chercheur dans sa composition ?
 - b. Le « tout quantitatif » : l'étude uniquement quantitative des grands corpus est-elle souhaitable ?
 - c. Y a-t-il des limites épistémologiques au tout quantitatif ? Au tout qualitatif ?
 - d. Si les deux approches sont nécessaires, y a-t-il un ordre plus pertinent ?
 - e. Comment intégrer les études qualitatives dans une étude sur grands corpus ? Comment prendre en charge les faits micro ?
3. Questions épistémologiques sur les commodités de langage :
 - a. La distinction quantitatif/qualitatif est-elle pertinente (un décompte des pronoms par exemple n'a-t-il pas une dimension qualitative ?) ?
 - b. De même pour les distinctions micro/macro et analyse globale/analyse fine (une étude macro n'apporte-t-elle rien de micro ou de fin en termes de connaissances ?) ?
 - c. Comment nommer ces deux approches ? Quelle est la pertinence de la distinction analyse lexicale/analyse discursive ? N'y retrouve-t-on pas le clivage sciences sociales/analyse du discours ?
 - d. La spécificité de l'AD réside-t-elle dans l'approche qualitative et la prise en compte des présupposés linguistiques dans l'utilisation des logiciels ?

Une publication des articles sélectionnés est prévue dans une revue sous forme d'un numéro thématique.

Les propositions de 300 mots environ sont à envoyer **avant le 19 novembre 2011** à thierry.guilbert@u-picardie.fr

Retour des avis du comité scientifique : 21 janvier 2012

Adresses :

Université de Picardie – UFR des Lettres
Chemin du Thil 80025 Amiens cedex1

Centre universitaire de recherches sur l'action publique et le politique – Epistémologie et sciences sociales (CURAPP - UMR 6054)

Placette Lafleur

80000 Amiens

Quelques éléments bibliographiques :

- Brunet, É. (1978) *Le Vocabulaire de Jean Giraudoux, structure et évolution. Statistique et informatique appliquées à l'étude des textes à partir des données du Trésor de la Langue Française*. Genève: Slatkine.
- Brunet, É. (2007) Le corpus comme une boule. In F. Rastier et M. Ballabriga (éds), *Corpus en lettres et sciences sociales. Des documents numériques à l'interprétation*. PUM : Toulouse.
- Culioli, A. (1968) La formalisation en linguistique. *Cahiers pour l'Analyse*, pp. 106-117.
- Fiala, P. (1994) L'interprétation en lexicométrie. Une approche quantitative des données lexicales. *Langue française* 103, pp. 113-122.
- Ghiglione, R., Matalon, B. et Bacri, N. (1985) *Les Dires analysés : l'analyse propositionnelle du discours*. Presses universitaires de Vincennes.
- Guilhaumou, J. (2006) *Discours et événement : l'histoire langagière des concepts*, Besançon : Presses universitaires de Franche-Comté.
- Guilhaumou, J. (1986) L'historien du discours et la lexicométrie. *Histoire et Mesure* 1(3-4), pp. 27-46.
- Labbé, D. (1990) *Le vocabulaire de François Mitterrand*. Paris: Presses de Sciences Po.
- Lafon, P. (1984) *Dépouillement et statistique en lexicométrie*. Kliensieck.
- Lebart, L. et Salem, A. (1994): *Statistique textuelle*. Paris: Dunod.
- Léon, J. (2010) ADD69 : archéologie d'une étrange machine. In Paveau M.-A. (éd), *La théorie du discours. Fragments d'histoire et de critique*, *Semen* 29, pp. 89-109.
- Mazière, F. (2005) *L'analyse du discours*. Paris : PUF.
- Mazière, F. & Guilhaumou, J. (2010) Ainsi nous qui sommes dans le Mississippi. In Paveau M.-A. (éd), *La théorie du discours. Fragments d'histoire et de critique*, *Semen* 29, pp. 69-88.
- Mayaffre, D. (2002) L'Herméneutique numérique, *L'Astrolabe* www.uottawa.ca/academic/arts/astrolabe/articles/art0031.htm
- Mayaffre, D. (2004) *Paroles de Président, Jacques Chirac et le discours présidentiel sous la Ve République*. Paris: Honoré Champion.
- Mayaffre, D. (2005) Les corpus politiques : Objet, méthode et contenu. Introduction. *Corpus* 4.
- Muller, C. (1973) *Initiation aux méthodes de la statistique linguistique*, Champion.
- Pêcheux, M. (1969) *Analyse automatique du discours*. Paris: Dunod.
- Pêcheux, M. (1975) *Les vérités de la Palice : linguistique, sémantique, philosophie*. Paris : Maspero.
- Robin, R. (1986) Postface. L'Analyse du Discours entre la linguistique et les sciences humaines : l'éternel malentendu. *Langages* 81, 121-128
- Salem, A. (1993) De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical. *Mots - Les langages du politiques* 36, pp. 74-83.
- Tournier, M. (1975) *Un vocabulaire ouvrier en 1848. Essai de lexicométrie*. Saint-Cloud: Publication de l'École Normale Supérieure.